

AD _____

Award Number: DAMD17-00-1-0538

TITLE: DNA repair and checkpoint genes as NF1 modifiers

PRINCIPAL INVESTIGATOR: Andre Bernards, Ph.D.

CONTRACTING ORGANIZATION: Massachusetts General Hospital
Boston, Massachusetts 02114

REPORT DATE: October 2002

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20030313 120

REPORT DOCUMENTATION PAGEForm Approved
OMB No. 074-0188

maintaining
the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)**2. REPORT DATE**
October 2002**3. REPORT TYPE AND DATES COVERED**

Annual (1 October 2001 - 30 September 2002)

4. TITLE AND SUBTITLE

DNA repair and checkpoint genes as NF1 modifiers

5. FUNDING NUMBER

DAMD17-00-1-0538

6. AUTHOR(S)

Andre Bernards, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)Massachusetts General Hospital
Boston, Massachusetts 02114
email - abernarde@helix.mgh.harvard.edu**8. PERFORMING ORGANIZATION
REPORT NUMBER****9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012**10. SPONSORING / MONITORING
AGENCY REPORT NUMBER****11. SUPPLEMENTARY NOTES****12a. DISTRIBUTION / AVAILABILITY STATEMENT**

Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE**13. ABSTRACT (Maximum 200 Words)**

This study aims to determine whether common protein altering SNP alleles of DNA repair or DNA damage-associated checkpoint genes are associated with higher or lower than average neurofibroma burden. To date we identified 746 protein altering alleles of 273 candidate modifier genes and we prioritized 138 common SNPs for analysis. We also established contacts with 252 NF1 patients and enrolled 66 eligible individuals. Since for several reasons recruitment was running behind schedule, we enlisted additional clinical collaborators. Contingent upon obtaining regulatory approval, DNA from 459 eligible patients is currently available for analysis. With recruitment continuing, we evaluated various high throughput genotyping methods. We identified three robust methods which we have used to determine >10,000 individual genotypes for 23 SNPs in 19 genes. By contrast, we found single base extension fluorescence polarization genotyping to be less reliable. Three grant proposals have been submitted based on preliminary data obtained in this project, and one NIH R01 grant was recently awarded.

14. Subject Terms (keywords previously assigned to proposal abstract or terms which apply to this award)

NF1 modifier genes; DNA repair genes; Cell cycle checkpoint genes; Single nucleotide polymorphisms; Genotyping assays; Protein truncation assays

15. NUMBER OF PAGES

12

16. PRICE CODE**17. SECURITY CLASSIFICATION**OF REPORT Unclassified
Unclassified**18. SECURITY CLASSIFICATION**OF THIS PAGE Unclassified
Unclassified**19. SECURITY CLASSIFICATION**OF ABSTRACT
Unclassified**20. LIMITATION OF ABSTRACT**

Unlimited

Table of Contents

Cover.....	1
SF 298.....	2
Table of Contents.....	3
Introduction	4
Body.....	4-10
Key Research Accomplishments.....	10
Reportable Outcomes.....	10-11
Conclusions.....	11-12
References.....	12
Appendices.....	n/a

Introduction

Neurofibromatosis type 1 (NF1) is a common genetic disorder that affects 2 to 3 per 10,000 worldwide. Patients are at increased risk of developing diverse symptoms, the most common of which include skin pigmentation defects, benign tumors associated with the peripheral nervous system, termed neurofibromas, and learning problems (Huson and Hughes, 1994). NF1 is paradigmatic for a disease with variable expressivity and genetic studies have implicated symptom-specific modifier genes as important determinants of clinical severity in NF1 (Easton et al., 1993; Szudek et al., 2000). This project aims at creating the resources to identify genetic modifiers of neurofibroma burden and to explore whether genes involved in maintaining genome stability play rate limiting roles in neurofibroma development. We focus on genes that modify neurofibroma development because these benign tumors contribute significantly to the overall morbidity of NF1 and because their numerical variability is a cause for significant patient anxiety as well as a major problem for clinical trials. Moreover, modifier genes are believed to play an important role in determining neurofibroma burden.

Body

The Statement of Work listed as **Task 1** the creation of computerized patient and modifier gene databases. This task was accomplished as planned during the first month of funding, but we have continued to modify and expand the single nucleotide polymorphism (SNP) database far beyond what we had envisaged during year 2. The patient database includes names, sex, dates of birth, clinical information (neurofibroma numbers), contact information, details about consent procedures, summaries of email messages and other contacts, codes used to identify samples in the laboratory, and other information if available. It currently contains information on 252 NF1 patients. Among these patients, 20 were seen at MGH by our collaborator Dr. Mia MacCollin. An additional 27 patients were brought to our attention by collaborating with Dr Andreas Kurtz, as suggested by the integration panel. The remaining 205 patients contacted the Principal Investigator or the project associated Genetic Counselor after learning about this study, mostly from notices posted by patient organizations.

Our proposal was to perform a case-control allele association study among 300 to 600 NF1 patients who represent the top and bottom 20% of neurofibroma burden. We proposed to genotype common protein altering SNP alleles of candidate modifier genes identified in a screen performed by collaborating researchers at the MIT Center for Genome Research. In practice, the MIT screen only scanned a small number of the candidate modifier genes identified by us. Thus, rather than limiting ourselves to the few genes analyzed at MIT, we invested considerable "data mining" effort to identify candidate modifier alleles among a comprehensive set of genes implicated in maintaining genome stability. This far more ambitious approach was made possible by the identification of well over one million SNPs during the early phases of the human genome project. As noted in our previous annual report, mining of online SNP and literature databases during the first year of funding identified 325 protein altering SNPs in 185 potential neurofibroma modifiers. 57 of these missense SNPs (17.5%) had a minor allele frequency $>4\%$. Continued data mining has presently identified 746 nonsynonymous alleles of 273 potential genome stability genes. The genes that we have analyzed include 20 genes implicated in base excision repair, 10 disease genes associated with increased sensitivity to DNA damage, 14 genes related to DNA damage response genes from other species, 16 DNA polymerase subunits, 7 DNA replication checkpoint genes, 16 genes involved in homologous recombination, 11 mismatch excision repair genes, 17 mitotic spindle checkpoint genes, 10 genes involved in nonhomologous end joining, 31 nucleotide excision repair genes, 9 genes involved in post-replication repair, 41 genes with a suspected DNA repair function, and 84 genes in various other categories. Among the latter group are several potential breast cancer susceptibility modifiers, which were included because BRCA1 and BRCA2 have roles in DNA repair and because in the absence of a fully assembled NF1 patient DNA panel, we practiced high throughput SNP genotyping using available somatic DNAs from 274 early onset (diagnosis before 40 years of age) breast cancer patients and a similar number of controls (FitzGerald et al., 1997). We obtained separate funding from the Avon Corporation to support this related project. Among the 746 missense SNPs identified thus far, 138 (18.5%) have a reported variant allele frequency $>4\%$, 148 have an allele frequency between 1 and 4%, 185 are in the $<1\%$ allele frequency class, and for 275 SNPs the allele frequency remains unknown. As noted before, we are most interested in SNPs in the $>4\%$ allele frequency category, since less common SNPs are unlikely to produce statistically significant results given the size of our patient panel.

Although public databases such as dbSNP or GeneSNP continue to improve, data quality still leaves much to be desired (Marsh et al., 2002). Thus, a large proportion of database entries still represent SNPs identified exclusively *in silico*, for example by comparing EST sequence traces. Typically, no allele frequencies are known for such SNPs and their reality remains in doubt. Online databases also remain subject to frequent and unpredictable change, and for many genes SNPs are listed without information as to whether they affect protein sequence. For all genes in our database we manually identified nonsynonymous SNPs. This is a time consuming process, but storing the maps used to identify SNPs as part of each gene's database record makes the evaluation of future SNP updates straightforward. For typical SNPs, our database lists minor allele frequency, the sequence around the polymorphism, information on whether the SNP affects evolutionary conserved amino acids (determined by performing BLASTP searches; SNPs that alter evolutionary conserved amino acids will be analyzed with highest priority), details about genotyping methods

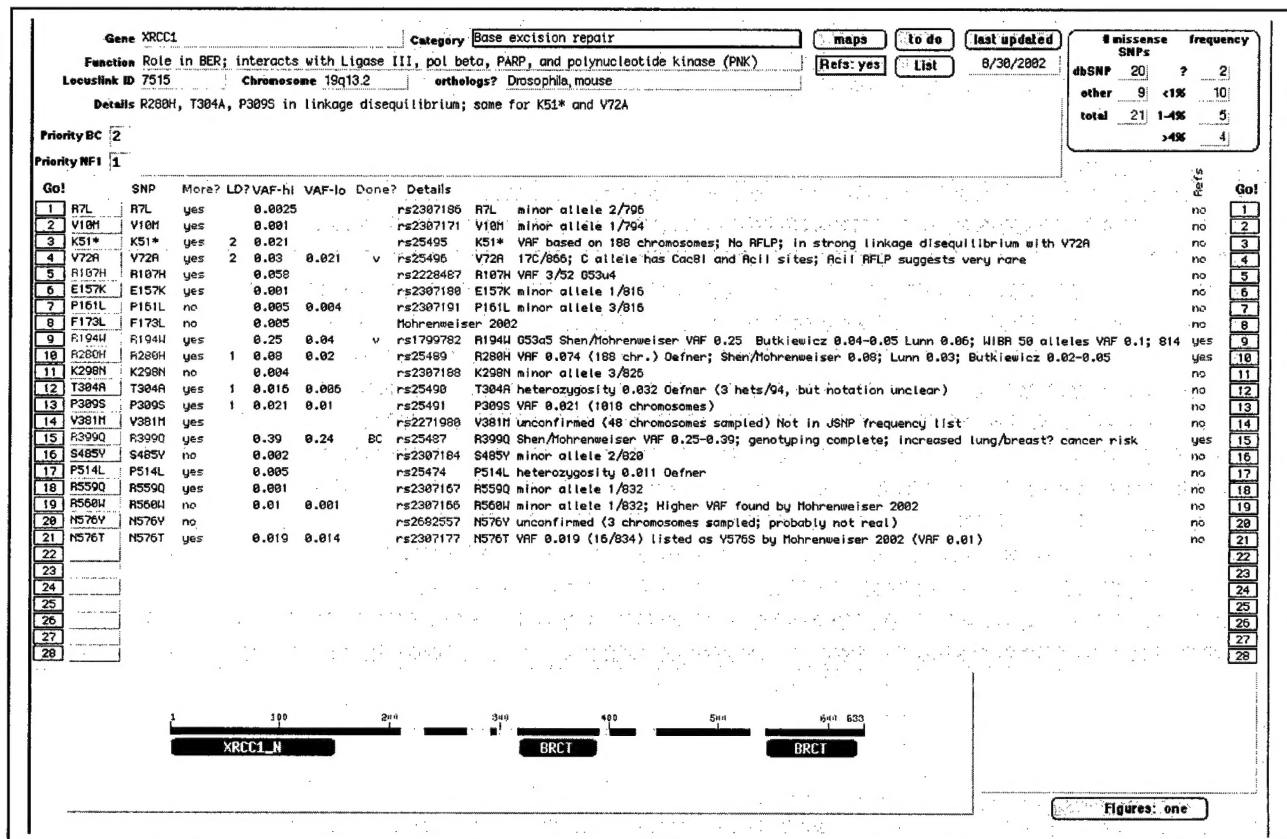


Figure 1. Main layout of SNP survey database. Relevant details are discussed in the text.

(PCR primer design, etc), and abstracts of papers that cite the SNP. The database also includes a computer generated domain structure of each protein, which helps to identify SNPs in potentially important protein segments. An important detail is that our overall database (current size 20.7 MB) consist of two integrated relational databases with gene-specific or SNP-specific information. Figure 1 shows the main SNP database layout for the *XRCC1* base excision repair gene.

Anticipating the need to efficiently process and analyze bulk genotyping data, during the past year we designed a separate genotyping results database. This database centrally stores genotyping data, output files from the Analyst-AD fluorescence polarization plate reader, or scanned gel pictures for SNPs genotyped by restriction fragment length polymorphism (RFLP) or allele specific PCR (ASP) methods. Importantly, the results database automatically calculates several basic statistical and other parameters from entered genotype data. Thus, entering observed genotypes calculates allele frequencies among cases and controls, expected allele frequencies based on Hardy-Weinberg equilibrium, χ^2 *P* values for observed allele distributions assuming both recessive and dominant models, and odds ratios with 95% confidence intervals for all genotypes. Having a database that performs these basic calculations does not substitute for more sophisticated biostatistical analysis, but is invaluable in practice.

Beyond creating the required bioinformatics resources, much of the remainder of this project was contingent upon our ability to recruit 300 eligible NF1 patients within 15 months and up to 600 eligible patients within two years. Thus, **Task 2** involved the analysis of a limited number of MIT discovered missense SNPs in peripheral blood DNA samples from 150 high and 150 low neurofibroma number patients during months 1-15, while **Task 3** was to confirm any detected allele association in 300 additional high and low neurofibroma burden patients during the remaining nine months. **Task 4** was to perform protein truncation assays to detect additional loss-of-function mutations among genes that showed positive allele associations. Soon after the start of this project it became apparent that our recruitment goals, based on estimates provided by clinical collaborators, had been unrealistic. Thus, Dr. Korf at Boston Children's Hospital had estimated to contribute between 60 and 100 patients annually, and Dr. MacCollin at MGH had indicated she would contribute between 40 and 50 eligible patients each year. The remaining patients were to be recruited by advertising this study nationally.

At this time we have enrolled 66 patients by means of our original recruitment strategy. Thus, we enrolled 44 of 204 patients who contacted us in response to various notices, 7 or 27 patients that were brought to our attention by Dr. Kurtz, 14 of 18 patients deemed eligible by Dr. MacCollin, and 1 of 1 patient referred to us by Dr. Korf. Obviously, recruitment from all sources has run far behind schedule. Among important reasons for this shortfall is that Dr. Korf gave up his directorship of the Boston Children's Hospital NF clinic just before the start of this project. We also did not anticipate that the Army IRB would not allow the recruitment of patients younger than 18 years of age, which excluded the majority of patients seen at this clinic. Another problem was that Dr. MacCollin went without a clinical coordinator for nine month and has only recently begun to contribute patients. Among the 204 patients who contacted us directly, 105 eligible patients have so far received consent and blood drawing kits, but only 44 have returned consent forms and blood samples so far.

Our original recruitment plan relied too heavily on the enthusiastic participation of two local NF clinics. Another problem was that patients recruited outside of these clinics would not be clinically evaluated, but rather would be recruited based on self-reported neurofibroma numbers. We sought to remedy both problems by enlisting additional clinical collaborators. However, all domestic clinicians approached by us balked at participating in an Army funded study given the burdensome regulatory process. We had more success enlisting collaborators in Europe and Table 1, taken from a recent grant application, lists six clinicians who have agreed to recruit eligible patients for this project.

Collaborator	Location	# DNAs available	# prospective patients
Evans, Gareth	Manchester, UK	0	150
Ferner, Rosalie	London, UK	0	>100
Lázaro, Conxi	Barcelona, Spain	55	30-60
Legius, Eric	Leuven, Belgium	0	>75
Mautner, Victor-Felix	Hamburg, Germany	288	300
Messiaen, Ludwine	Ghent, Belgium	50	50-70
Locally recruited	Boston, MA	66	100
Total		459	805-875

Table 1; Clinical collaborators and number of available or to-be-recruited eligible patients.

The collaborators listed in Table 1 have DNAs from 393 eligible patients available for analysis. Beyond this number, they expect to recruit around 700 more patients within three years (in order to have additional statistical power to detect associations, in our recent grant proposals we increased the patient panel size from 600 to 1200 differentially affected patients). It is important to note that those listed in Table 1 have only agreed to participate if contributing patients anonymously circumvents the need for obtaining separate Army IRB approval. As a test case, we recently amended our human studies protocol to allow the analysis of 55 patient DNAs provided to us without identifying information by Dr. Conxi Lázaro. During the ensuing three month long comprehensive re-review of our entire protocol all patient recruitment was suspended, which contributed to the low number of patients recruited during the past year. However, the fact that we did eventually obtain regulatory approval suggests that no fundamental problems stand in the way of this approach. Thus, although patient recruitment has been more problematic than anticipated, we do expect to achieve our original recruitment goals in the near future.

Our proposal was to genotype a limited number of missense SNPs discovered at MIT using a single base extension fluorescence resonance energy transfer (SBE-FRET) protocol. However, before the start of this project our collaborators at MIT had replaced SBE-FRET by a lower cost single base extension fluorescence polarization (SBE-FP) protocol. In this homogenous method SNP containing DNA segments are PCR amplified, followed by enzymatic degradation of primers and nucleotides, and extension of an unlabeled primer that abuts the SNP with fluorescent chain terminators. Incorporation of either one or both chain terminators is measured as an increase in fluorescence polarization (Kwok, 2002). In our first annual report we noted that our original plan to use MIT Genome Center equipment to read SNP genotypes turned out to be unworkable and that we had acquired our own LJI-Analyst-AD 96/484 well fluorescence polarization plate reader. After spending considerable effort optimizing and evaluating the reliability of SBE-FP genotyping, we have reluctantly concluded that SBE-FP genotyping is not as problem-free as suggested. Thus, rather than close to 100% successful assays and >99% accuracy with little optimization (Hsu et al., 2001), only about 70% of our assays work and in typical cases accuracy is only around 95%. We arrived at these numbers by genotyping multiple SNPs in parallel by SBE-FP and RFLP or ASP methods. Using a combination of all three methods, we successfully determined 10038 individual genotypes for 23 SNPs in 19 different genes during the past year.

Current SNP genotyping methods remain cumbersome and costly (typically \$0.50 to \$1,30 per genotype), making the analysis of candidate modifier SNPs the only practical approach. Although there is much excitement about matrix-assisted laser desorption-ionization time of flight (MALDI-TOF) mass spectroscopy based SNP genotyping, at \$0.60 per four-fold multiplexed assay this method also remains far too costly for anything but candidate SNP screens. However, methods to simultaneously genotype thousands of SNPs as pennies per genotype are on the horizon, suggesting the feasibility of less biased, genome wide SNP association studies. We are particularly interested in a microarray-based method which is projected to allow, within 12 months or so, the genotyping of 100,000 SNPs at around \$0.01 per SNP using 0.5 µg of reduced complexity genomic DNA as a probe. Thus, we envisage that the patient DNA panel assembled during this pilot project may eventually be used for comprehensive genome-wide SNP haplotype determinations. We recently submitted grant proposals to the NIH and the Army NF Research Program to support this work. The requested one year no cost extension for this project would allow us to continue patient recruitment while these new grant proposals are being considered.

Key Research Accomplishments

1. Designed and implemented patient information database
2. Designed and implemented Genome Stability Gene SNP database
3. Contacted 251 NF1 patients and enrolled 66.
4. Identified clinical collaborators that will contribute >1000 additional patients
5. Determined a total of 10038 individual genotypes for 24 SNPs in 19 genes while evaluating SBE-FP and other genotyping methods.

Reportable Outcomes

- Meeting abstract. Analysis of *NF1* function in *Drosophila*. Anna Tchoudakova, James Walker, Peter McKenney, Iswar Hariharan and André Bernards. NNFF International Consortium for the molecular biology of NF1 and NF2. Aspen, CO. June 8-12, 2002.
- Patient database, Genome Stability Gene SNP database listing information on 746 missense SNPs in 273 candidate genome stability genes, and SNP Genotype Analysis database.

- NIH R01 Grant Application. Title: Quantitative Phenotyping and Genotype-Phenotype Correlations in NF1; Principal Investigator: Bruce R. Korf. Results from the current project were used as preliminary data in this recently awarded application, which uses a discordant sib pair strategy to perform intrafamilial and interfamilial comparisons of dermal neurofibroma and café-au-lait macule numbers for identification of modifier loci.
- NIH R01 Grant Application. Title: Studies of neurofibromatosis-1 modifier genes. Principal Investigator: Andre Bernards. Results from the current project were used as preliminary data in this application, whose main aims include allele association studies to evaluate three classes of potential neurofibroma burden modifiers
- Army NF Research Program Investigator-Initiated Research Proposal. Title: Studies of neurofibromatosis-1 modifier genes. Principal Investigator: Andre Bernards. Results from the current project were used as preliminary data in this application, which has complete scientific and budgetary overlap with the NIH application listed above.

Conclusions

The main goals of this 2 year project were to collect somatic DNAs from 600 NF1 patients that represent the top and bottom 20% of neurofibroma burden and to use this resource to evaluate whether protein-altering alleles of genes implicated in maintaining genome stability are associated with a high or low neurofibroma burden. We encountered several significant problems during the execution of this project. Firstly, our plan to genotype missense SNPs in candidate modifier genes identified in a SNP discovery screen at MIT ran into problems when it became apparent that only a small fraction of candidate modifier genes had been analyzed in the MIT screen. This required us to perform time consuming data mining in order to identify a comprehensive set of candidate modifier alleles. Secondly, our plan to use MIT Genome Center equipment to read SNP genotypes turned out to be impractical, requiring us to buy our own Analyst-AD 96/384 well fluorescence polarization plate reader. Thirdly, notwithstanding published reports to the contrary, in our hands SBE-FP genotyping is not reliable enough and requires too much optimization to allow efficient high throughput genotyping of multiple SNPs. While RFLP- ASP-, or Pyrosequencing-based methods are more robust in our experience, these procedures remain too costly or too labor intensive for true high throughput genotyping. We are currently evaluating mass spectroscopy-

based genotyping, and we are in discussions with researchers at Affymetrix about their soon-to-be-launched microarray-based SNP genotyping platform. Thus, although we have not yet reached our stated goals, the experience gained during this two year pilot project has been invaluable and has allowed us to submit grant proposals that aim to continue and significantly expand our efforts to identify modifiers of neurofibroma development in NF1.

References

- Easton, D. F., Ponder, M. A., Huson, S. M., and Ponder, B. A. (1993). An analysis of variation in expression of neurofibromatosis (NF) type 1 (NF1): evidence for modifying genes. *Am J Hum Genet* 53, 305-313.
- FitzGerald, M. G., Bean, J. M., Hegde, S. R., Unsal, H., MacDonald, D. J., Harkin, D. P., Finkelstein, D. M., Isselbacher, K. J., and Haber, D. A. (1997). Heterozygous ATM mutations do not contribute to early onset of breast cancer. *Nat Genet* 15, 307-310.
- Hsu, T. M., Chen, X., Duan, S., Miller, R. D., and Kwok, P. Y. (2001). Universal SNP genotyping assay with fluorescence polarization detection. *Biotechniques* 31, 560, 562, 564-568
- Huson, S. M., and Hughes, R. A. C., eds. (1994). *The neurofibromatoses: A pathogenetic and clinical overview*, First edn (London, Chapman & Hall Medical).
- Kwok, P. Y. (2002). SNP genotyping with fluorescence polarization detection. *Hum Mutat* 19, 315-323.
- Marsh, S., Kwok, P., and McLeod, H. L. (2002). SNP databases and pharmacogenetics: A great start but a long way to go. *Hum Mutation* 20, 174-179.
- Szudek, J., Birch, P., Riccardi, V. M., Evans, D. G., and Friedman, J. M. (2000). Associations of clinical features in neurofibromatosis 1 (NF1). *Genet Epidemiol* 19, 429-439.

Appendices

None